

**METHOD OF FABRICATING AN INTEGRATED CIRCUIT WITH A DIELECTRIC
LAYER EXPOSED TO A HYDROGEN-BEARING NITROGEN SOURCE**

Claim of Priority

[0001] This application is a continuation of U.S. Patent Application No. 10/378,568, filed March 3, 2003, which is a divisional of U.S. Patent Application No. 09/544,505, filed April 6, 2000, now U.S. Patent No. 6,559,007 issued May 6, 2003, which is incorporated in its entirety by reference herein. This application is related to U.S. Patent Application No. 10/378,354, filed March 3, 2003 and U. S. Patent Application No. 10/378,573, filed March 3, 2003.

Field of the Invention

[0002] The present invention generally relates to transistor gate dielectrics and methods of fabricating the same. More particularly, the invention relates to processes and structures for improving tunnel oxide quality in erasable programmable read-only memories (EEPROMs).

Background of the Invention

[0003] Memory devices such as erasable programmable read-only memories (EPROMs), electrically erasable programmable read-only memories (EEPROMs), or flash erasable programmable read-only memories (FEPROMs) are erasable and reusable memory cells which are used in digital cellular phones, digital cameras, LAN switches, cards for notebook computers, etc. A memory cell operates by storing electric charge (representing an "on" state) on an electrically isolated floating gate, which is incorporated into a transistor. This stored charge affects the behavior of the transistor, thereby providing a way to read the memory element. It is therefore crucial that the memory cell be able to maintain the stored charge over time, so that charge leakage does not cause data errors by converting "on" states to "off."

[0004] A memory cell typically consists of a transistor, a floating gate, and a control gate above the floating gate in a stacked gate structure. The floating gate, typically composed of polysilicon, is electrically isolated from the underlying semiconductor substrate by a thin dielectric layer, which is typically formed of silicon oxide. Because charge is transferred across the dielectric layer by quantum-mechanical tunneling, this dielectric layer is often referred to as a "tunnel oxide" layer. Such tunnel oxide layers are typically approximately 100 Å thick. Properties of the tunnel oxide must be strictly controlled to ensure the ability to read and write by tunneling, while avoiding

data loss through charge leakage. The control gate is positioned above the floating gate, and is electrically isolated from the floating gate by a storage dielectric layer, such as oxide-nitride-oxide (ONO). Electrical access to the floating gate is therefore only through capacitors.

[0005] Storing charge on the floating gate programs a memory cell. This is achieved via hot-electron injection by applying a high positive voltage (approximately 12 V) to the control gate, and a high drain-to-source bias voltage (approximately 6 V). An inversion region is created between the source and drain by the control gate voltage, and electrons are accelerated from the source to the drain by the drain bias voltage. Some fraction of these electrons will have sufficient energy to surmount the tunnel oxide barrier height and reach the floating gate. The floating gate is therefore programmed by collecting and storing these electrons to represent an “on” state.

[0006] An EPROM device can be erased (*i.e.*, returned to an “off” state) by exposing the floating gate to ultraviolet light, which excites the stored electrons out of the floating gate. The erasure of an EEPROM or FEPRM cell is accomplished via Fowler-Nordheim tunneling, in which an electric field is sufficient for the stored electrons to traverse the tunnel oxide and enter the substrate, thereby reducing the stored charge in the floating gate. Under this mechanism for discharging the floating gate, a large negative voltage (e.g., -10 V) is applied to the control gate, and a positive voltage (e.g., 5-6 V) is applied to the source while the drain is left floating. Electrons then tunnel from the floating gate through the tunnel oxide, and are accelerated into the source. Because both the programming and erasing of a memory element takes place via charge transfer processes across the tunnel oxide layer, it is important to minimize the density of interface states and other defects in the form of charge traps in this region which would otherwise create a mechanism for charge leakage through the tunnel oxide.

[0007] Current memory devices have shown improvements in data retention by the formation of a thin oxynitride layer, typically formed by exposure of the thermal oxide to either NO or N₂O at elevated temperatures. The thin nitrided interface improves the tunnel oxide’s resistance to the creation of defects by electrons during the write and erase sequence. Another possible method of reducing charge leakage is to remove species such as hydroxyl (OH) from the tunnel oxide prior to final encapsulation. Nitridation also forms a barrier against further degradation by impurity migration. The density of charge trapping sites due to OH in the tunnel oxide region is thereby significantly reduced, resulting in devices with improved data retention and reliability.

[0008] Once the stacked gate structure has been fabricated and etched to the appropriate dimensions, the stacked gate structure is encapsulated in a liner layer, followed by the formation of an insulating layer, typically composed of thick, planarized borophosphosilicate glass (BPSG). The liner layer between the source/drain regions and the BPSG, often composed of a low pressure chemical vapor deposition (LPCVD) oxide, serves to minimize out-diffusion of contaminants and dopants from the BPSG. Such out-diffusion might otherwise affect the performance of underlying devices.

[0009] Additional improvements in memory device performance have been achieved by performing "alloy" steps later in the fabrication process. In an alloy step, the integrated circuit being fabricated is exposed to hydrogen while being annealed at low temperatures (less than 450°C). Incorporation of hydrogen is believed to tie up dangling bonds at the tunnel oxide region, thereby eliminating interface states that would otherwise contribute to charge trapping and device degradation. The positive effects of alloy steps are well known in the art. Unfortunately, hydrogen annealing and/or OH exposure at high temperatures can be detrimental to transistor devices. See, e.g., Vanheusden et al., "Positive Charging of Buried SiO₂ by Hydrogenation," APPL. PHYS. LETT., Vol. 64, No. 19 (May 9, 1994), pp. 2575-77. Typical oxide liners are not effective barriers against H₂ or OH. Accordingly, high temperature processing allows mobile OH molecules to diffuse into the tunnel oxide after formation, degrading data retention of the resultant part.

[0010] While processes have been developed to improve gate dielectric quality, as measured by improved data retention in flash memory devices, for example, a finite soft error rate remains. Accordingly a need exists for further improvements in the fabrication of transistor gate dielectrics.

Summary of the Invention

[0011] In accordance with one aspect of the present invention, a method is provided for fabricating an integrated circuit. The method includes forming a dielectric layer on a semiconductor surface and exposing the dielectric layer to a hydrogen-bearing nitrogen source. A conductive layer is then deposited directly over the dielectric layer. After patterning to define a gate stack, a conformal insulating liner is formed over the gate stack. The liner incorporates nitrogen.

[0012] In accordance with another aspect of the invention, an integrated circuit is provided with a plurality of transistors. Each transistor includes a hydrogen-passivated gate

dielectric layer, a stacked gate structure above the gate dielectric layer, and a conformal liner layer covering the stacked gate structure. The liner layer includes nitrogen.

[0013] In the illustrated embodiments, a tunnel oxide in an EEPROM device is nitrified with ammonia. Hydrogen is thus incorporated into the tunnel oxide, along with nitrogen. The gate stack is etched and completed, including protective sidewall spacers and dielectric cap, and the stack lined with a silicon nitride or oxynitride.

[0014] Advantageously, the liner serves as an excellent diffusion barrier against contamination of the gate stack from hydrogen or OH, particularly protecting the tunnel oxide. While such effective diffusion barriers also inhibit hydrogen diffusion from subsequent alloy steps, meant for occupying dangling bonds at the substrate and tunnel oxide interface, the prior ammonia nitridation compensates for this hydrogen blocking by providing hydrogen to the oxide prior to sealing the gate stack with the liner.

Brief Description of the Drawings

[0015] Figure 1 is a flow chart, generally illustrating a process flow in accordance with a preferred embodiment of the present invention.

[0016] Figure 2 illustrates a device structure encompassing one preferred embodiment of the present invention.

Detailed Description of the Preferred Embodiment

[0017] While illustrated in the context of an electrically erasable programmable read only memory (EEPROM) device for flash memory circuits, persons skilled in the art will readily find application for the present invention to fabrication of other semiconductor integrated circuit devices. In particular, methods disclosed herein are applicable to improving the gate dielectrics in a wide variety of transistor designs with a wide variety of process flows. The methods described herein, however, have particular utility for improving the performance of tunnel dielectrics.

[0018] Figure 1 is a flow chart which generally illustrates a process flow in accordance with one preferred embodiment of the present invention, and Figure 2 illustrates an EEPROM transistor 10 in a flash memory cell, constructed in accordance with this preferred embodiment. In the following description of the preferred embodiments, the named process flow steps are found in Figure 1 and the numbered structural elements refer to Figure 2. It will be understood, however, that elements may differ in appearance during fabrication as compared to the illustrated final structure.

For example, the gate stack layers described below are preferably blanket deposited upon one another prior to definition of the gate electrode by photolithography and etch.

[0019] The EEPROM transistor 10 produced by the preferred embodiment of the present invention is fabricated over a semiconductor substrate 20, which includes doped regions corresponding to a source 22 and a drain 24. In the illustrated embodiment, the substrate 20 comprises the upper portion of a single-crystal silicon wafer. In general, however, the substrate can comprise any semiconductor structure or layer in which the lowest level of integrated electrical devices are formed. As shown, a transistor gate structure 26 overlies the substrate 20, particularly over a channel region of the substrate 20 between the source 22 and the drain 24. The active areas 22, 24 and transistor gate 26 together define the EEPROM transistor 10.

[0020] In the preferred embodiment of the present invention, the first step in the fabrication of the stacked gate structure 26, as illustrated in Figure 1, is the formation 110 of an initial dielectric layer 30 on the surface of the substrate 20. The dielectric 30 preferably comprises an oxide, and more particularly silicon oxide, though the skilled artisan will appreciate that the present invention will have utility in conjunction with other types of oxide. An exemplary alternative oxide comprises tantalum pentoxide (Ta_2O_5). In the illustrated embodiment, formation 100 of the dielectric 30 comprises thermal oxidation of the substrate surface, but persons skilled in the art are able to select an appropriate method of forming the tunnel dielectric layer 30 from the various possible methods. The thickness of the illustrated dielectric layer 30 is preferably between about 10 Å and 300 Å, more preferably between about 30 Å and 150 Å, and most preferably between about 80 Å and 115 Å (*e.g.*, for a flash memory part).

[0021] The next step shown in Figure 1 comprises nitridation 110 of the the initial dielectric layer 30. At least a top surface 40 thereby incorporates nitrogen and forms an effective diffusion barrier. The amount of nitrogen incorporated into the oxide preferably forms a barrier to diffusion of hydrogen species out of the dielectric. The nitridation 110 can be performed either during the growth or in a subsequent anneal of the initial dielectric layer 30.

[0022] In the preferred embodiment of the present invention, the nitridation 110 is performed by exposing the initial dielectric layer 30 to a hydrogen-bearing nitrogen source, particularly ammonia vapor (NH_3), while the substrate 20 is held at elevated temperatures. The amount of nitrogen and hydrogen incorporated in the dielectric increases with increasing substrate

temperature during exposure to NH_3 . For example, a substrate temperature between about 850°C and 950°C incorporates approximately 2% - 3% into the resultant tunnel dielectric 30, 40.

[0023] Too little or too much nitrogen and/or hydrogen can adversely affect device performance. Preferably, nitrogen and hydrogen are incorporated without nitrogen- or hydrogen-induced defects. Conditions are arranged to incorporate between about 0.1 atomic % and 5.0 atomic % nitrogen in the tunnel dielectric 30, 40, more preferably between about 0.1 atomic % and 0.6 atomic %, and most preferably between about 0.3 atomic % and 0.6 atomic %. The resultant tunnel dielectric of the illustrated embodiment comprises silicon oxynitride, or SiO_xN_y .

[0024] In particular, in order to reduce the impact of trap sites created by exposure of oxide to hydrogen at higher temperatures, the substrate temperature is desirably kept below about 850°C during nitridation 110, more preferably in the range of about 600°C to 850°C, and most preferably about 650°C to 750°C. The NH_3 exposure during nitridation 110 is preferably maintained for between about 1 second and 60 minutes, more preferably between about 30 seconds and 300 seconds. Exemplary pressure during NH_3 nitridation 110, for the most preferred temperature range above, is about 300-760 Torr.

[0025] Subsequent to nitridation 110, the layers of the transistor gate stack can be completed 120 in any suitable fashion, typically entailing numerous processing steps. In the illustrated embodiment, a floating gate 50 is formed by depositing a conductive layer (preferably doped polysilicon) onto the tunnel dielectric 30, 40. Doping of the polysilicon of the floating gate 50 can be either *in situ* (i.e., while the floating gate 50 is being formed) or it can be a separate step after the polysilicon deposition. Persons skilled in the art are able to select appropriate materials and methods for creating the floating gate 50 with a particular set of characteristics.

[0026] After the formation of the floating gate 50, the formation of the gate stack continues by forming a storage dielectric layer 60 on the floating gate 50. In the illustrated embodiment, the storage dielectric layer 60 is composed of oxide-nitride-oxide (ONO), formed by methods known in the art. In other arrangements, high dielectric materials can be employed to improve capacitance of the EEPROM device. Persons skilled in the art can readily select appropriate materials and methods for creating the storage dielectric layer 60 for particular circuit designs.

[0027] The stack completion 120 continues by then forming a control gate 70 over the storage dielectric layer 60. The control gate 70 can be composed of various conductive materials,

including, but not limited to, polysilicon, metal and/or metal silicide. Upon formation of the control gate 70, a cap insulator layer 80 is preferably formed, comprising an insulator such as silicon nitride or silicon oxide, over the control gate 70.

[0028] After the completion 120 of the gate stack, the stack is patterned 130, such as by conventional photolithography and etch processes, to define the gate electrode structure 26. In the illustrated embodiment, spacers 90a and 90b are also formed 140 along the sidewalls of the stacked gate structure 26. Conventional blanket deposition of an insulating material followed by directional spacer etch can be employed for spacer formation 140.

[0029] In the preferred embodiment of the present invention, as illustrated in Figure 1, the gate 26 and other surrounding areas are then covered 150 by a substantially conformal liner layer 92. While conventional oxides effectively block dopant diffusion, such as boron and phosphorus from overlying BPSG, the liner 92 of the preferred embodiments also forms a good diffusion barrier against hydroxyl and hydrogen species. The liner 92 comprises an insulating material, preferably incorporating both silicon and nitrogen. Thus, preferred liner materials includes silicon nitride, silicon oxynitride or a multiple layer laminate including one or both of nitride and oxynitride. The skilled artisan will readily appreciate that the invention can comprise any of a number of other suitable materials that form effective barriers to diffusion of OH and hydrogen.

[0030] The illustrated liner 92 comprises silicon oxynitride ($\text{Si}_x\text{N}_y\text{O}_z$), where z is in the range of about zero to 0.3 (30 atomic %), more preferably between about 0.1 and 0.2 and y is greater than zero, more preferably between about 0.3 and 0.57 and most preferably between about 0.4 and 0.57. Exemplary ratios include $\text{Si}_{0.43}\text{N}_{0.57}\text{O}_0$ (Si_3N_4), $\text{Si}_{0.5}\text{N}_{0.4}\text{O}_{0.12}$, $\text{Si}_{0.45}\text{N}_{0.35}\text{O}_{0.2}$, etc.

[0031] The liner 92 can be formed 150 by any suitable manner, but is preferably formed by chemical vapor deposition (CVD) to ensure good step coverage over the topography of the patterned gate electrodes 26 across the substrate. CVD processes are well developed for each of Si_3N_4 and silicon oxynitride. For example, a silicon source (*e.g.*, SiH_2Cl_2), nitrogen source (*e.g.*, NH_3) and oxygen source (*e.g.*, O_2 , N_2O) can be reacted in a CVD process. Varying the relative source gas flow or partial pressure can vary the nitrogen content, as will be appreciated by the skilled artisan. Alternatively, CVD oxide can be nitridized, *e.g.*, by remote plasma nitridation, to incorporate 1-40 atomic % nitrogen, more preferably 4-15 atomic %.

[0032] Subsequent to forming the liner layer 92 in the preferred embodiment of the present invention, an interlevel insulating layer 94 is deposited 160 over the structure. Typically composed of BPSG, the layer 94 serves to electrically isolate underlying devices, such as the EEPROM transistor 10, from overlying interconnects. Accordingly, the insulating layer 94 is preferably between about 6,000 Å and 20,000 Å in thickness.

[0033] After forming the liner, state of the art integrated circuit fabrication typically employs an alloy step 170, in which the structure is exposed to a hydrogen-containing ambient at low temperatures (less than about 500°C). In the illustrated embodiment, the liner 92 serves as a barrier to OH and H diffusion. As previously noted, such species can cause damage to the tunnel oxide 30 in the form of charge leakage paths. However, hydrogen already incorporated into the tunnel oxide provides passivation of substrate interface. The alloy step 170 is thus only optional, and can be omitted in the preferred embodiments. If an alloy is employed, the liner 92 advantageously prevents OH and other contamination from reaching the tunnel oxide 30.

[0034] After depositing 160 the interlayer dielectric layer 110, the integrated circuit is completed by additional fabrication steps. Typically, such steps include metallization processes 180, interconnecting various devices of the integrated circuit. In order to make contact electrical contact between metal layers and the electronic devices, holes or vias are etched through the interlevel dielectric layers between levels, such as the insulating layer 94, and then filled with conductors. Contact to the transistor control gate 70 and active areas in the substrate 20, for example, require contact through the insulating layer 94 and the liner 92.

[0035] Optional hydrogen anneals 190 can also be performed during or after metallization to further reduce the dangling bonds at the substrate surface, including the region of the tunnel dielectric layer 30, 40. Diffusion is facilitated somewhat through the contacts. As noted with respect to hydrogen anneal after deposition 160 of the insulating layer 94, hydrogen diffusion can also be enhanced by grinding the backside of the workpiece.

[0036] The integrated circuit is then completed by formation of bond pads and final passivation 200, such as by deposition of a further silicon oxynitride layer or other suitable passivation material. As will be appreciated by the skilled artisan, the passivation layer forms a seal against moisture or other corrosive agents.

[0037] The use of NH_3 during nitridation 110 of the initial dielectric layer 30 advantageously improves the resultant device performance. As will be appreciated by the skilled artisan, nitridation of the tunnel oxide with hydrogen-bearing species enables the use of a more effective diffusion barrier for the liner 92. The more effective liner 92 limits exposure of the tunnel oxide 40 to outgassing of OH and H during subsequent processing, which improves data retention in the resultant chip. Using NH_3 also advantageously provides hydrogen atoms within the tunnel dielectric, preferably resulting in a hydrogen content between about 0.1% and 10% in the tunnel oxide. Moreover, this hydrogen is effectively retained within the tunnel dielectric due to the presence of a nitrogen-rich surface on the tunnel dielectric, locking hydrogen in during subsequent processing. The hydrogen is then available at later stages for passivating dangling bonds at the substrate surface and in the region of the tunnel dielectric layer. Because the preferred embodiment of the present invention introduces hydrogen atoms during the formation of the oxynitride layer 40, the efficacy of any subsequent alloying steps is of less importance. Less aggressive hydrogen anneals can be employed. In one embodiment, subsequent hydrogen anneals are eliminated altogether. In another embodiment, only post-metallization hydrogen anneal is employed.

[0038] Furthermore, the use of NH_3 for nitridation permits the use of a more effective diffusion barrier for the liner layer 100 that protects the transistor gate stack. Whereas very effective diffusion barriers protect the transistor from water or OH contamination during BPSG deposition, for example, they also inhibit desirable hydrogen diffusion. The preferred embodiments, however, provide and seal hydrogen in the tunnel dielectric at an early stage, thereby facilitating use of more effective liner material without sacrificing hydrogen passivation of the tunnel dielectric and substrate interface. In the preferred embodiment of the present invention, the liner layer 100 comprises silicon nitride (Si_3N_4) or silicon oxynitride (SiO_xN_y). These nitrogen-containing compounds are more effective barriers to OH diffusion than, for example, LPCVD oxide.

[0039] Experiments have confirmed the effectiveness of the preferred ammonia rapid thermal anneal (RTA) upon device performance. In the absence of hydrogen alloy annealing, devices formed with ammonia RTA of tunnel oxide performed better than devices formed with longer and more expensive NO anneal of tunnel oxide at 800°C for 30 minutes, 800°C for 15 minutes and 850°C for 30 minutes. In particular, plots of drain current I_D against gate voltage V_G indicate a steeper slope (e.g., by a ratio of 39 to 34-37) for the ammonia anneal. The steeper slope

indicates greater carrier mobility in the channel. Accordingly, the ammonia anneal serves not only to harden the tunnel oxide, but also provides hydrogen passivation of dangling bonds at the substrate and gate dielectric interface.

[0040] Although described above in connection with particular embodiments of the present invention, it should be understood the descriptions of the embodiments are illustrative of the invention and are not intended to be limiting. Various modifications and applications may occur to those skilled in the art without departing from the true spirit and scope of the invention as defined in the appended claims.